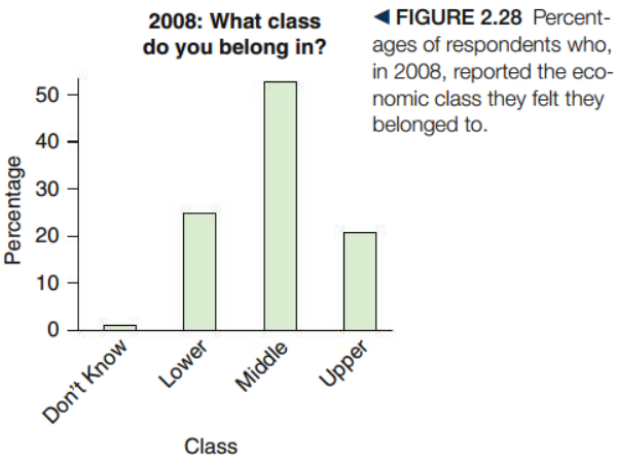
2.4: Summarizing Categorical Distributions

Definition.

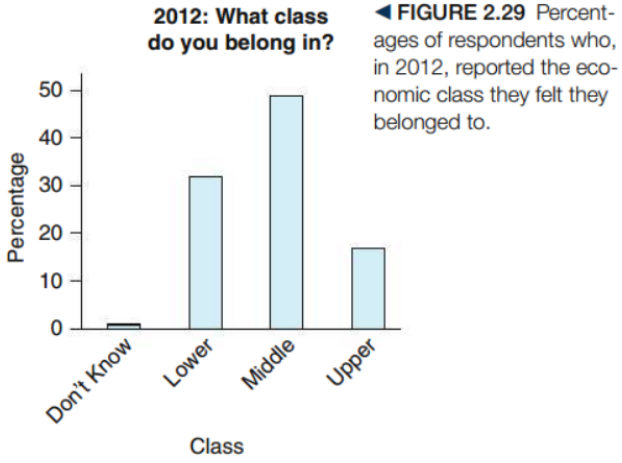
- The category that occurs most often is called the **mode** (similar to the usage with numerical variables).
- A distribution with a lot of *diversity* is said to have a high **variability**.

Note: A categorical variable is considered bimodal *only* if two categories are nearly tied for the mode.

Example. Below are the results of a survey conducted in 2008, and again in 2012. How do the response compare?



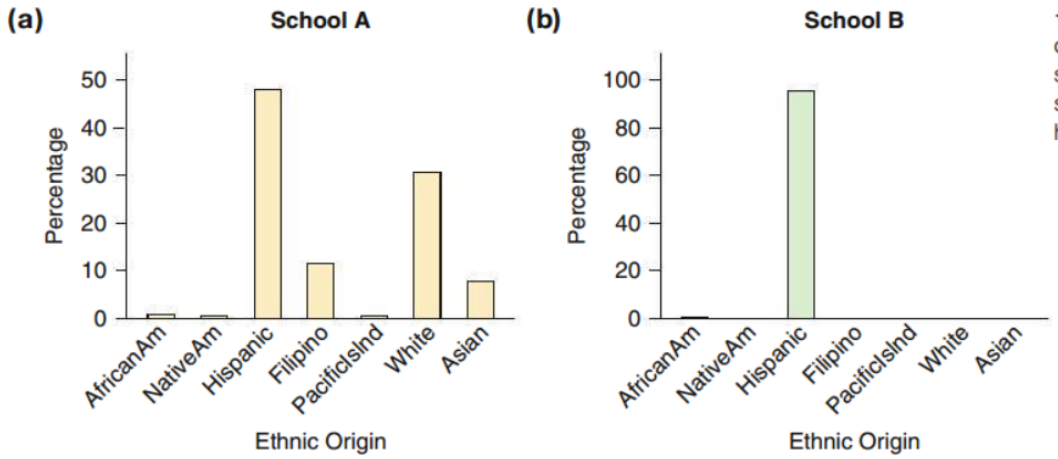
◀ **FIGURE 2.28** Percentages of respondents who, in 2008, reported the economic class they felt they belonged to.



◀ **FIGURE 2.29** Percentages of respondents who, in 2012, reported the economic class they felt they belonged to.

Mode is "Middle" in both 2008 and 2012
"Lower" has a higher percentage in 2012
Results from 2012 appear to have more variability

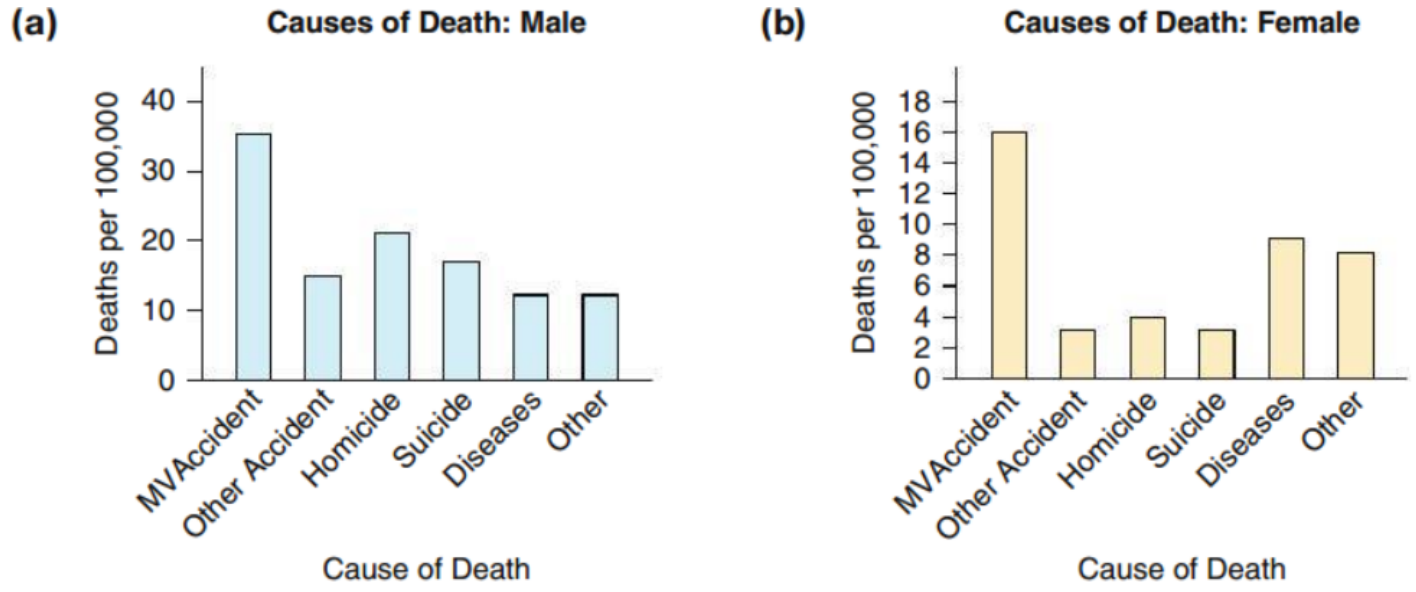
Example. The ethnic composition of two schools in the Los Angeles City School System is presented in the bar charts below. Which school has the greater variability in ethnicity?



◀ **FIGURE 2.30** Percentages of students at two Los Angeles schools who are identified with several ethnic groups. Which school has more ethnic variability?

More diversity in School A -> more variability

Example. Compare the distributions below. What is the mode for each graph? Which graph demonstrates more variability?

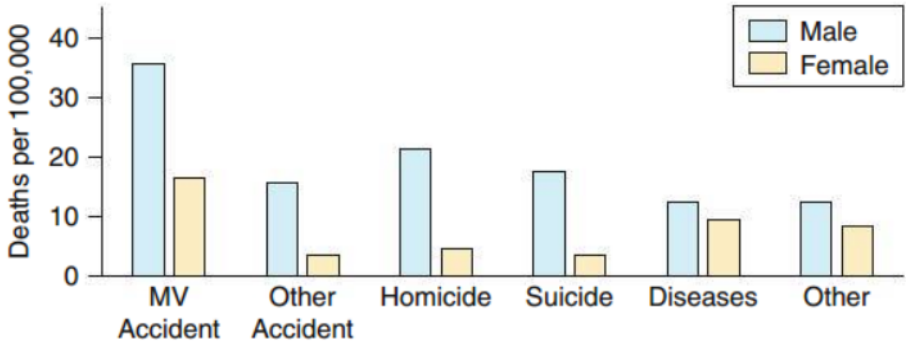


▲ **FIGURE 2.32** The number of deaths per 100,000 males **(a)** and females **(b)** for people 15 to 24 years old in a one-year period.

MVA are the mode for both males and females
* Frequency of females is 1/2 that of males
Males have more variability

Example. Compare the combined bar graphs below to the graphs above.

▶ **FIGURE 2.33** Death rates of males and females, graphed side by side.



2.5: Interpreting Graphs

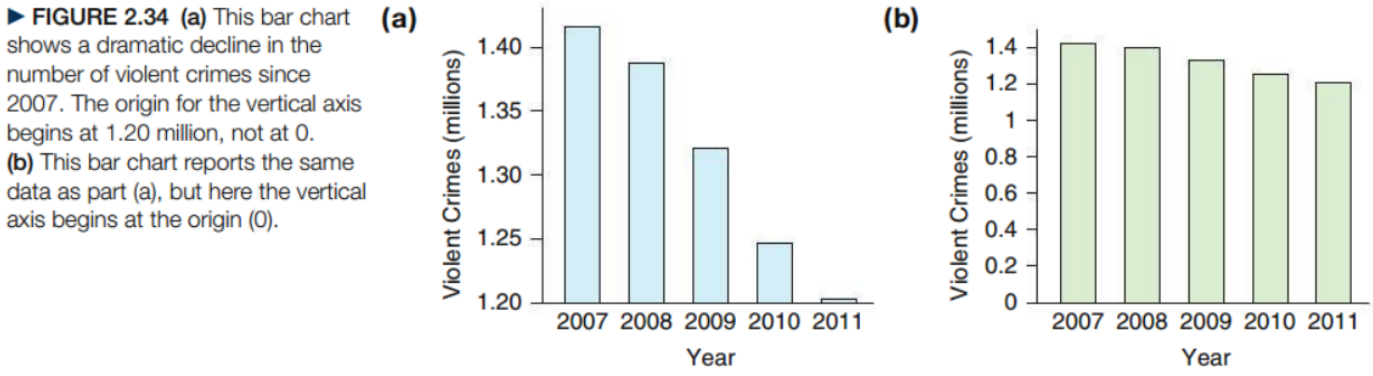
Appropriate Graphs:
The type of data determines the type of graph you should use

Numerical Data	Categorical Data
Dot plot	Pie chart
Histogram	Bar graph
Stem-and-leaf plot	

Appropriate Measures:
The type of data determines how the distribution of data should be described

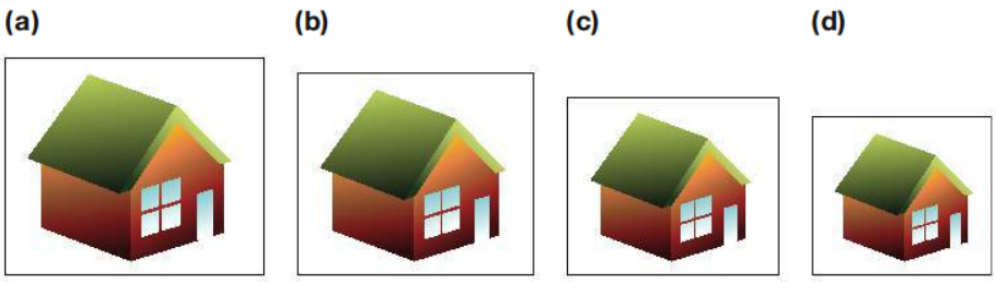
Numerical Data	Categorical Data
Shape	Mode
Center	—
Spread	Variability

Misleading Graphs: • Inappropriate scaling (starting at a nonzero value)

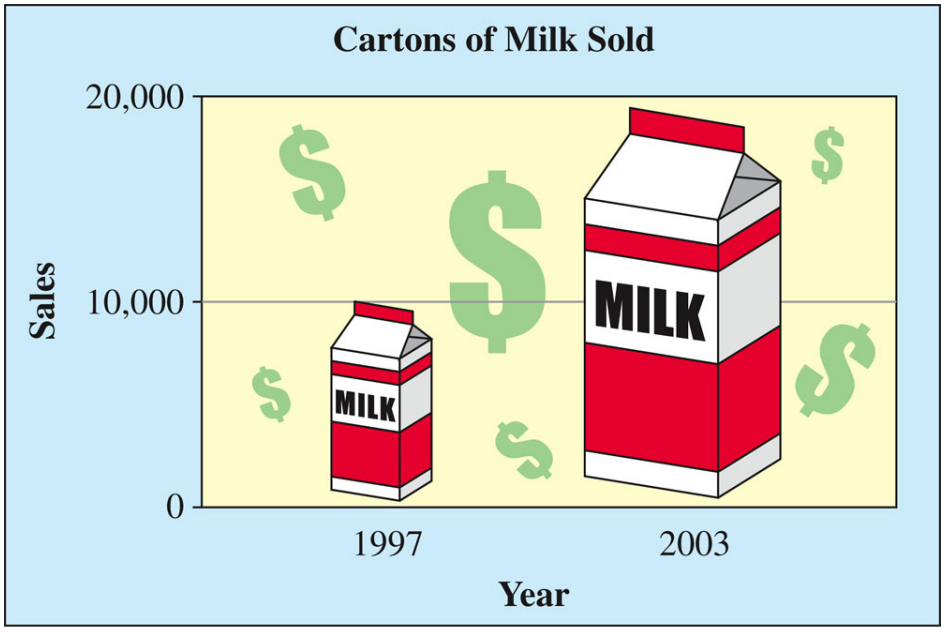


▶ **FIGURE 2.34** **(a)** This bar chart shows a dramatic decline in the number of violent crimes since 2007. The origin for the vertical axis begins at 1.20 million, not at 0. **(b)** This bar chart reports the same data as part (a), but here the vertical axis begins at the origin (0).

- Icons of different sizes instead of bars of proportionate heights:



◀ **FIGURE 2.35** Deceptive graphs: Image **(a)** represents 7.1 million homes sold in 2005, image **(b)** represents 6.5 million homes sold in 2006, image **(c)** represents 5.8 million homes sold in 2007, and image **(d)** represents 4.9 million homes sold in 2008. (Source: *L.A. Times*, April 30, 2008)



- Avoid the use of 3D graphs: